

Interacció Document-Humà a través d'interfícies de paper augmentat

Jordi Arús Garrigós

Resum— Els sistemes d'Interacció Humà-Document s'han creat per a intentar unir els avantatges del món digital i del món físic. En aquest projecte es vol intentar determinar si resultaria atractiva per als usuaris una aplicació desenvolupada per a un d'aquests sistemes, concretament el sistema DoD desenvolupat pel CVC. S'ha desenvolupat una aplicació que permet seleccionar text d'un document, prèviament introduït a la base de dades, i fer una traducció d'aquest text, i una recerca a Internet de termes relacionats. Un cop l'aplicació es va desenvolupar, es va provar en un entorn real per recollir l'opinió dels usuaris. Aquests van trobar que seria una aplicació útil, tot i que van trobar necessàries millores i funcionalitats addicionals, com més opcions de traducció o la possibilitat de poder utilitzar qualsevol document, introduint una llibreria OCR a l'aplicació. De totes les propostes recollides s'ha extret un seguit de vies de millora per si es segueix treballant en aquest tipus de sistemes.

Paraules clau— Interacció Humà-Document, projecció, paper augmentat, traducció, recerca

Abstract— Human-Document Interaction Systems have been created to try to unite the perks of the digital world and the physical world. The goal of this project is to try and determine if an application developed for one of these systems, specifically the DoD system developed by the CVC, would be attractive to users. An application has been developed, and it allows users to select text from a document, previously inserted in the database, and get a translation of this text, and also a related terms search. Once the application was finished, it was tested on a real environment to get feedback from the users. They found that this application would be useful, although it lacked some improvements and additional functions, such as having more translation options or the possibility of using any kind of document, by introducing an OCR library to the application. Out of all the proposals, a few improvement lines have been extracted in the event that work continues on these kind of systems.

Index Terms— Human-Document Interaction, projection, augmented paper, translation, search



1 INTRODUCCIÓ

RECENTMENT s'ha vist com cada cop més les persones prefereixen utilitzar plataformes digitals, com podrien ser ordinadors, tauletes o smartphones, per a l'edició i visualització ràpida de documents. Però donat que els documents en paper ofereixen una sèrie d'avantatges i comoditats per als usuaris que serien difícils de replicar en suports digitals, i en processos cognitius, és més preferible buscar una manera d'aprofitar els beneficis dels dos suports. Per això s'han ideat sistemes d'interacció Humà-Document, que mitjançant diferents aplicacions ofereixen diverses maneres d'interacció amb diferents tipus de documents. Una de les motivacions en estudiar aquests sistemes d'interacció Humà-Document seria, en paraules del professor Fernando Vilariño seria “enriquir els documents físics utilitzant documents digitals (vídeos, imatges, ...)”. Una altra motivació seria, com hem comentat abans, buscar una interacció fàcil i intuïtiva entre documents físics i documents digitals en entorns reals que poguessin facilitar la realització de certes tas-

ques aprofitant els avantatges dels dos dominis.

Si bé fa temps que es busquen alternatives d'interacció fluida entre el domini físic i diferents tipus d'interfícies digitals, hi ha diferents propostes desenvolupades per diferents equips d'investigació. D'entre els sistemes d'interacció Humà-Document[9] que existeixen, l'aplicació a desenvolupar en aquest treball es basarà en el sistema DoD[1], desenvolupat per investigadors del Centre de Visió per Computador. D'aquest sistema tota la posta a punt i la interconnexió entre els diferents mòduls ja estaven prèviament desenvolupats, per tant la idea de treball sobre el sistema és crear una nova aplicació per a generar noves maneres d'interacció entre suports físics en paper i interfícies digitals, tot oferint una sèrie de noves funcionalitats a l'usuari.

En aquest article es detallarà tot el procés seguit durant tot el desenvolupament d'aquest Treball de Final de Grau. Primer de tot es definirà l'abast i els objectius del projecte, es farà una breu explicació de l'estat de l'art en sistemes d'Interacció Humà-Document, i seguidament es detallarà amb quina metodologia s'ha treballat, i quin procés s'ha seguit en el desenvolupament. Finalment, es

-
- E-mail de contacte: jordi.arus@e-campus.uab.cat
 - Menció realitzada: Computació
 - Treball tutoritzat per: Dimosthenis Karatzas, CVC
 - Curs 2015/16

presentaran els resultats obtinguts i d'aquests se n'extrauran una sèrie de conclusions.

2 OBJECTIUS

Amb aquest Treball de Final de Grau volem donar resposta a la següent pregunta: Resulta atractiva per a un usuari una aplicació desenvolupada utilitzant una interfície d'Interacció Humà-Document? Són realment funcionals aquests tipus de sistemes?

La manera de validar aquesta hipòtesi ha sigut desenvolupar una aplicació utilitzant un sistema d'interacció Humà-Document que permeti a un usuari, introduint un document físic en el sistema, utilitzar una interfície que li permeti realitzar funcionalitats addicionals sobre els continguts del document.

També es van definir una sèrie d'objectius addicionals, que són els següents.

1. Dissenyar noves interfícies basades en el paper per accedir a funcionalitats digitals. En concret, dissenyar una interfície interactiva per accedir a funcionalitats de traducció i recerca Web utilitzant el paper com a mitjà d'interacció
2. Desenvolupar una aplicació de prova de la nova interfície
3. Avaluar l'aplicació en un entorn real amb usuaris finals

A més es va definir un objectiu opcional, que es duria a terme al finalitzar la resta dels objectius si hi havia temps disponible:

- Modificar l'aplicació de forma progressiva, afegint funcionalitats o millorant les ja existents

Tots aquests objectius addicionals s'han complert, i l'objectiu opcional no s'ha pogut dur a terme per falta de temps en la realització del projecte.

3 ESTAT DE L'ART

3.1 Inicis

Fa dècades que els investigadors treballen sistemes d'interacció Humà-Document. Un dels exemples més primerencs seria el treball de Wellner amb el seu Digital Desk [2], que consisteix en tecnologia de projecció utilitzada juntament amb càmeres que permeten reconèixer el document que es troba sobre l'escriptori[7]. Si bé disposava d'algunes aplicacions funcionals, com poder seleccionar números d'un document per a introduir-los a una aplicació calculadora, també plantejava diferents escenaris sobre com aquesta tecnologia podria avançar en un futur, com per exemple prendre notes seleccionant una part d'un document i seguidament escriure-hi mitjançant un teclat.

3.2 Avenços posteriors

Des d'aquest primer acostament no es va avançar substancialment fins fa pocs anys, amb projectes com Lumi-

nAR[3] del MIT, i la tecnologia portable SixthSense[4] que també utilitzen tecnologia de projector i reconeixement gestual, tot i que utilitzen conceptes bastant diferents. El LuminAR combina la projecció i càmera amb tecnologia robòtica, permetent més flexibilitat en quant al posicionament de la superfície de projecció/interacció. En canvi, la tecnologia SixthSense està més enfocada a ser una tecnologia similar al que avui en dia s'anomenen wearables[13], en el qual la càmera i el projector les subjecta l'usuari. Això teòricament permetria, doncs, que es pogués interactuar en pràcticament qualsevol superfície que l'usuari trobi a la seva disposició. Un altre prototipus interessant és HideOut[11] de Disney, que mitjançant un projector portable i documents amb marcadors especials amb tinta especial infraroja permet que, tot i moure el projector al voltant del document, aquest sempre sigui localitzat correctament a través de marcadors ocults, i poder així projectar la imatge correctament des de gairebé qualsevol angle. Finalment, el projecte StripTic[12] utilitza informació obtinguda de tires de paper per a ajudar en el control del tràfic aeri.

D'entre les diferents aproximacions de sistemes d'interacció Humà-Document, la utilitzada per al projecte serà el sistema DoD[1], desenvolupat per investigadors del Centre de Visió per Computador. A grans trets, és un sistema que utilitza tecnologia de projector i reconeixement gestual i d'imatges mitjançant càmeres. Una de les diferències d'aquest sistema respecte d'altres ja mencionats és la manera d'interactuar[8]. A part del reconeixement de gestos manuals, es permet també l'ús de punters infrarojos. L'avantatge d'això respecte a l'ús de bolígrafs digitals és que pot treballar sobre qualsevol part de la superfície, i no és necessària una preparació especial prèvia del document. A més, alguns dels sistemes mencionats anteriorment estableixen l'enllaç entre el document físic i el digital utilitzant marcadors, bé visibles o infrarojos, o bé altres sistemes basats en registrar el contingut, que poden presentar problemes aplicant-los a documents en que el text sigui escàs o no segueixi convencions de text en llenguatges llatins (orientació de l'escriptura, ordre de lectura, ...). En canvi, el sistema DoD utilitza reconeixement de característiques a nivell de píxel, utilitzant concretament mètodes SIFT[10], que no presentaria per tant aquests problemes.

3.3 Altres alternatives

Un dels casos en els que s'ha indagat és la realitat augmentada. Si bé aquest sistema són d'interès per l'anàlisi d'imatges en temps real i pel fet que poden enriquir documents físics aportant més informació, presenten certes limitacions. Necessàriament estan lligats a un dispositiu amb una pantalla on es visualitza la informació, com pot ser un smartphone o una tablet, i mentre s'utilitza, l'usuari no pot interactuar amb el document físic ja que està agafant el dispositiu. Per tant, en comptes de dispositius de realitat augmentada com aquests, es va optar per solucions de realitat augmentada projectada, com el que s'utilitzarà per aquest treball.

4 METODOLOGIA

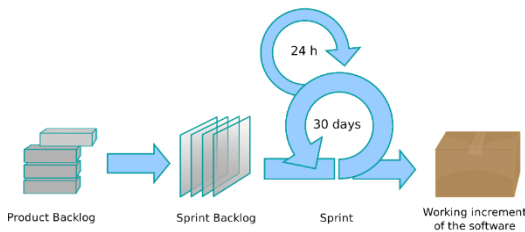


Figura 1. Gràfic amb el funcionament general de Scrum

Per al desenvolupament de l'aplicació s'ha utilitzat una metodologia similar a la metodologia de desenvolupament àgil Scrum. És similar en el fet de que s'han definit una sèrie amb una durada de dues setmanes, i s'han anat definint les tasques a realitzar en cada iteració per a poder implementar una nova funcionalitat o tenir una nova versió de l'aplicació. S'ha triat aquesta manera de treballar ja que d'aquesta manera es poden anar definint metes a curt termini per anar avançant en el desenvolupament de l'aplicació, així com per aprofitar les reunions setmanals acordades amb el tutor de treball que marcaven el final d'una iteració i el principi de la següent.

En les reunions normalment participaven els següents actors. El tutor del treball, Dimosthenis Karatzas, en qualitat de Product Owner, l'alumne Jordi Arús com a desenvolupador del projecte, i, ocasionalment, alguns membres del Centre de Visió per Computador com Marc Castelló o Jordi Abella, en qualitat d'assistents tècnics.

Pel seguiment de cada sprint s'ha utilitzat l'eina Trello, tant a nivell personal de l'alumne com per a que el tutor, que també hi tenia accés, tingués una vista general de les tasques realitzades i les pendents.



Figura 2. Cronograma de la planificació indicant diferents milestones

En total s'han dut a terme 9 sprints, durant els quals es van definir com a milestones les diferents entregues de documents que s'han hagut de dur a terme. Durant aquests sprints s'anaven desenvolupant una sèrie de tasques, algunes ja planificades inicialment, i algunes sorgides d'afegir alguna nova funcionalitat. Es va fer, però, una divisió en diferents fases del procés que englobaven aquestes tasques. Aquestes fases són les següents:

1. Anàlisi del sistema sobre el qual es desenvoluparà el projecte
2. Desenvolupament de la versió preliminar

3. Milliores en el funcionament de l'aplicació (fase opcional)

Seguint el cronograma, la primera fase es va completar durant els primers 2 sprints. La segona fase es va allargar fins pràcticament l'entrega del segon informe de progress, és a dir, fins que no es va tenir finalitzada l'aplicació. La tercera fase està lligada a l'objectiu opcional que es va definir, i per tant també es va definir com a una fase opcional.

5 DESENVOLUPAMENT

5.1 Detalls tècnics del sistema DoD

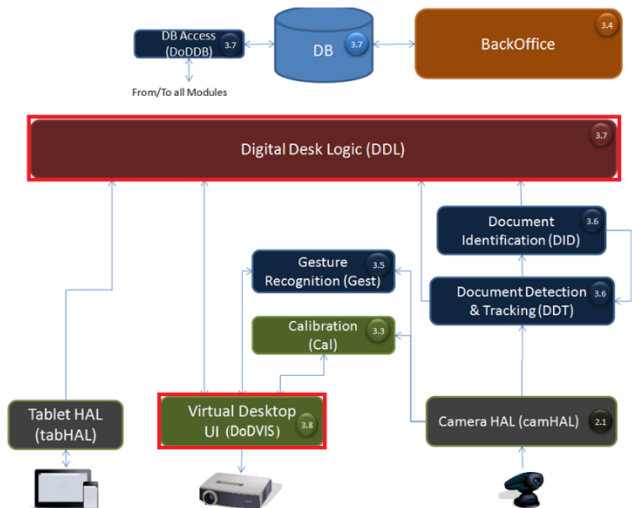


Figura 3. Esquema dels mòduls del sistema DoD

El sistema DoD consta d'una sèrie de mòduls independents que es comuniquen entre ells. Cadascun d'ells té una sèrie de funcions concretes, i alguns d'ells poden oferir accés a aquestes des d'altres mòduls. Els mòduls que s'han modificat durant aquest projecte, assenyalats a la Figura 3 anterior, són el mòdul de visualització (DoDVIS) i el mòdul de lògica (DDL). El mòdul DDL conté tota la lògica de l'aplicació, per tant és el mòdul principal d'aquest projecte, i del mòdul DoDVIS es va modificar la manera de representar les Hot Areas, ja que no estava correctament implementada. A més a més, s'han hagut d'introduir dades a la base de dades en algun punt. També s'ha interactuat amb els altres mòduls que conformen el sistema, però no se n'ha modificat el codi. El sistema està programat utilitzant el llenguatge de programació C++. Per a la interconnexió entre els diferents mòduls s'utilitza la llibreria Protocol Buffers per a la definició i encapsulat dels diferents missatges que es transmetran, i la llibreria ZeroMQ per a fer la comunicació pròpiament dita.

En quant a la configuració física, el sistema consta un projector amb tecnologia DLP, ja que ofereixen bona qualitat a un preu raonable. També consta de dues càmeres, una d'elles amb un filtre infraroig, per a capturar les imatges i la interacció amb els punters. El setup complet del sistema fixe costa aproximadament uns 3.000€. Si bé

existeix una instal·lació al Library Living Lab, on s'ha dut a terme una prova amb usuaris, i també una versió del sistema muntat sobre un suport de làmpada per a que sigui portable [1], es treballarà principalment sobre el sistema ubicat al Centre de Visió per Computador.

A part del sistema DoD, s'han utilitzat una sèrie de tecnologies i eines addicionals. Tot el codi del projecte es troba en un projecte de Visual Studio. El gestor de base de dades utilitzat és Oracle Express, i per a administrar-ne les dades s'ha utilitzat el programa PL-SQL Developer. Per a la definició de les àrees d'interès d'una manera més fàcil s'ha fet servir la plataforma TextExtraction [6] proporcionada pel CVC, que retorna la informació en fitxers XML. Es va desenvolupar un script en Python per a extreure fàcilment l'extracció de la informació dels fitxers XML i generar sentències SQL per a inserir els registres de forma correcta a la base de dades.

Per a les traduccions es fan consultes a la Microsoft Translate API, i per a la recerca de termes s'utilitzarà l'API DuckDuckGo. Totes dues utilitzen el mètode de comunicació REST, per tant s'ha implementat en el codi les funcions necessàries per a fer les consultes amb l'estructura necessària. Per a això van ser necessàries dues llibreries, SFML, que ja estava inclosa en el projecte degut a que s'utilitzava en altres mòduls, i cURL. També es va necessitar amb la llibreria cURL utilitzar la llibreria OpenSSL per a fer peticions segures.

En quant al disseny de la interfície, el sistema oferia una sèrie d'elements de visualització per a poder generar una interfície. Aquest eren etiquetes (labels), menús i submenús, i llistes d'elements (Listbox). Per tant, el disseny de la interfície va lligat a aquestes possibilitats que ens oferia el sistema. Es va decidir utilitzar els menús i llistes per a mostrar les opcions disponibles i que l'usuari les pogués seleccionar, i les etiquetes per mostrar informació i resultats, sense més interacció que la possibilitat de fer scroll pel text.

5.2 Fase inicial: anàlisi del sistema

En aquesta primera fase, es van produir les reunions inicials amb els professors, i és on es va realitzar el treball de documentació prèvia sobre les diferents eines, programari i llibreries que s'utilitzarien. Durant aquesta fase també es va poder tenir accés al manual del sistema DoD[5], per a posterior consulta durant el desenvolupament. També durant aquesta fase es va obtenir un projecte de Visual Studio en C++ sobre el qual es treballaria. Per a familiaritzar-se amb algunes de les funcions, especialment referides a l'intercanvi de missatges entre els diferents mòduls i la creació d'elements de la interfície, es va crear un mock-up amb una primera aproximació al que podria ser una interfície per utilitzar en l'aplicació final.

A continuació es van haver d'escollir els documents que s'utilitzarien per a mostrar el funcionament de l'aplicació. Després d'una recerca, es va escollir utilitzar el llibre *Introduction to Sociology*, obtingut de Open Stax College, concretament un data set de 60 pàgines.

L'aprenentatge es va dur a terme identificant punts característics de cada un dels documents, concretament es va definir que es fes l'aprenentatge seleccionant 1000 punts característics (keypoints) de cada document. Per a això es va utilitzar el mòdul d'identificació de documents (DID). Dels 60 documents introduïts en l'aprenentatge, tots van ser posteriorment correctament identificats pel sistema.

5.3 Segona fase: Desenvolupament de la versió preliminar

El primer que es va programar va ser la interfície i el flux d'interacció amb aquesta. Si bé hi havia una primera versió, aquesta no era suficient per diversos motius. Primerament, la interacció no era del tot intuïtiva en primera instància. A més, posteriorment aquesta interfície hauria de posicionar-se i moure's en relació amb la posició i moviment del document que posarem en el sistema.

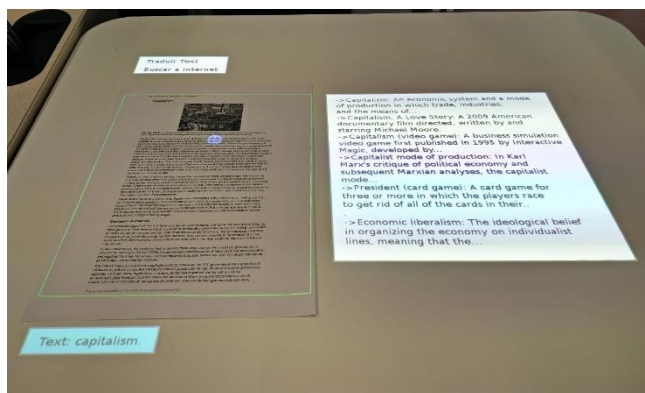


Figura 4. Mostra de l'aplicació i la seva interfície

Es van afegir també elements addicionals per a millorar l'experiència d'usuari, com per exemple una etiqueta que mostri a l'usuari exactament quin és el text que té seleccionat. A més, aquesta etiqueta mostra en temps real la selecció en el moment en què s'està seleccionant text novament.

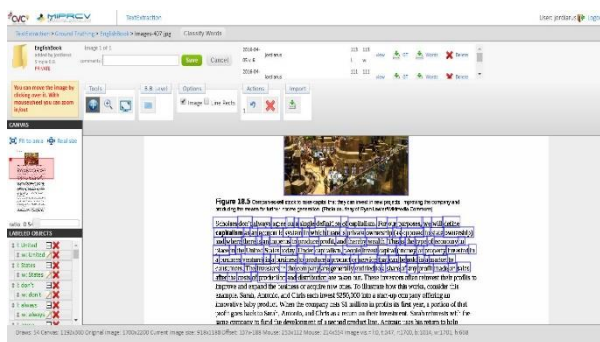


Figura 5. Interfície de l'aplicació de Ground Truthing

Seguidament es va començar a fer la definició de Hot Areas (o àrees d'interacció) en un document de prova. La idea era definir una Hot Area a cada paraula, ja que seran els elements que serviran en la nostra aplicació per a que l'usuari hi interactuï. Per a aquesta tasca s'ha utilitzat l'eina de Ground Truthing de la plataforma TextExtraction[6] proporcionada pel CVC, que retorna la informació

en fitxers XML. A més, es va crear un script en Python per a extreure fàcilment l'extracció de la informació dels fitxers XML i generar sentències SQL per a inserir els registres de forma correcta a la base de dades (DoDDB). Abans de definir les paraules del document, però, es va decidir introduir una sèrie d'àrees de control, per a intentar detectar possibles errors o modificacions a introduir tant en l'script com en la lògica de l'aplicació. De cada àrea la informació que tindrem concretament serà posició (coordenades relatives al document), mida (amplada i alçada, també relatives al document), i la transcripció.

Seguidament es va implementar el mètode de selecció de text. Si bé ja estava implementat des d'un principi el poder fer tap (clicar) sobre els objectes utilitzant un punter infraroig (mòdul de gestos GEST), per a seleccionar text es va decidir utilitzar una Bounding Box (caixa envoltant). La idea d'això seria definir, utilitzant dos punters, un rectangle, i seleccionar totes les paraules les coordenades de les quals es trobin dins d'aquest rectangle. D'aquestes àrees, en guardarem la transcripció i això constituirà el text seleccionat per l'usuari. Després amb aquest text es podran fer les operacions necessàries per a les funcionalitats que hem definit.

Les APIs utilitzades són les següents:

- Per a la traducció, Microsoft Translate API. En primera instància es volia utilitzar l'API de Google però, al descobrir que era de pagament, es va trobar aquesta alternativa que sí que ofereix serveis gratuïts, dins d'un límit mensual de 2 milions de caràcters
- Per a la recerca, DuckDuckGo. Similar al cas anterior, es va buscar una API gratuïta que ens tornés resultats de recerca, més concretament de termes relacionats amb el terme buscat

Finalment, es va haver de programar una funció que permetés fer consultes a les APIs utilitzades tant per la traducció com per a la recerca. Per a fer això s'ha utilitzat el mòdul Network de la llibreria SFML, llibreria de la qual s'utilitzaven altres mòduls en altres parts del sistema, com per exemple en la visualització d'elements gràfics. Aquesta funció s'ha fet parametritzada, de tal manera que permeti fer les peticions a qualsevol de les APIs que utilitzarem. Per exemple, en la funció de traduir text es definirà l'idioma origen (de moment serà només anglès), l'idioma de destí, que podrà ser català o castellà en primera instància, i s'afegirà a la URL on farem la petició aquesta informació acompanyada del text a traduir. Llavors cridarà a la funció de peticions HTTP que només s'encarregarà de retornar el resultat de la petició, i la funció de traducció seria l'encarregada de canviar el format del text de forma correcta per a ser mostrar per pantalla.

Cal destacar, però, que aquesta llibreria no va ser suficient. En el cas concret de l'API de Microsoft, requereix d'una autenticació prèvia per a utilitzar-la, per la qual s'ha de fer una petició segura mitjançant el protocol HTTPS, cosa que SFML no suporta, pel que es va haver d'utilitzar una llibreria addicional, a part de SFML. Es va

decidir utilitzar la llibreria cURL per a poder-ho fer. Amb aquesta llibreria es va fer una funció addicional per a fer una petició al servidor d'autenticació de Microsoft per a obtenir l'autorització per a poder fer posteriorment peticions de traducció. A més a més, a l'hora d'instal·lar la llibreria també es va necessitar utilitzar OpenSSL.

Un cop es van poder fer peticions de traducció, es va trobar un problema addicional. El sistema DoD, per com estan programats alguns dels seus mòduls, no mostra correctament caràcters especials com poden ser lletres accentuades, la lletra ç o la ll. Això és degut a la codificació del text, que no suporta caràcters Unicode (codificació UTF-8[15]). Per a solucionar això, el que s'hauria de fer implicaria modificar el codi de varis dels mòduls implicats en el procés de mostrar el text, com el mòdul de visualització. Aquesta modificació es podria fer en un futur com a possible millora de l'aplicació i del sistema DoD, però per l'aplicació preliminar que s'ha desenvolupat la solució triada ha estat processar el text traduït abans per a eliminar aquests caràcters especials i substituir-los per caràcters que sí que es puguin mostrar, com per exemple eliminar els accents de les lletres accentuades, substituir ç per c o ll per l.

Una altra de les funcionalitats que es volia implementar era el fet de que el text s'actualitzi a mesura que s'està duent a terme la selecció amb la bounding box. Per a fer això, però, va ser necessari utilitzar diferents homografies. Les coordenades de la bounding box obtingudes del missatge d'actualització (BoundingBoxUpdating) estan en coordenades de la càmera infraroja, mentre que les coordenades dels 4 cantons del document obtingudes es troben en coordenades de càmera visible. Per aquest motiu, es van haver de convertir aquestes coordenades en un mateix sistema de coordenades. En primera instància es va intentar convertir les coordenades de la bounding box de càmera infraroja a càmera visible, demanant al mòdul de calibració (CAL) la homografia corresponent, que és una funcionalitat que el mòdul permet. Es va veure, però, que aquest mòdul no ofereix homografies des de 0 cap a el sistema de coordenades de càmera infraroja, per tant es va haver de pensar una altra manera de fer-ho. Es va trobar que un dels fitxers del mòdul de lògica, que s'encarrega de dur a terme la calibració, sí que calcula la homografia des del sistema de coordenades de càmera infraroja cap al sistema de coordenades de workspace. Per tant, el que es va fer és treure d'aquí la homografia per la bounding box, i per a les coordenades del document demanar la homografia al mòdul de calibració, que sí que ens retornava correctament aquesta homografia.

Un cop obtingudes les dues homografies, les utilitzarem per a transformar els corresponents punts d'un sistema de coordenades a l'altre, i després calcular les coordenades en relació al document, per a poder utilitzar la mateixa funció que ja es va desenvolupar per a poder determinar quines de les paraules estem seleccionant amb la bounding box. D'aquesta manera, es podrà saber exactament quines paraules s'estan seleccionat mentre encara esti-

guem dibuixant la bounding box.

Una altra característica que es va afegir va ser que es mostrés un missatge indicant quan l'aplicació estava treballant. Donat que en algunes tasques, com la crida per fer traduccions o recerques, podria haver-hi un lleuger retard de mentre s'estava duent a terme l'operació en qüestió. Per tant es va triar de mostrar aquest missatge per a fer saber aquest fet a l'usuari.

6 MÈTODE D'AVUACIÓ DELS RESULTATS

Un cop finalitzada la implementació de totes aquestes funcionalitats, es van dur a terme totes les gestions per a poder fer una prova de l'aplicació en un entorn real. Per a fer això, l'aplicació es va deixar disponible al públic del Library Living Lab de la Biblioteca Miquel Batllori durant un període de temps, per a que provessin el seu funcionament i a més donessin el seu feedback, indicant què els ha agradat i què no, i recollint les seves possibles propostes de millora. Per tant, a part d'acordar amb el personal de la Biblioteca els horaris concrets per a poder mostrar l'aplicació, es van haver de confeccionar una sèrie de documents:

- Una "càpsula", d'estil similar a com podria ser un pòster d'un TFG com els que es poden veure a la facultat, per a explicar als usuaris la idea general del projecte i l'aplicació
- Un full d'instruccions, per a explicar de forma gràfica com els usuaris poden interactuar amb l'aplicació
- Un qüestionari/enquesta per a que els usuaris l'emplenin un cop hagin provat l'aplicació

L'enquesta consta de les següents preguntes:

1. Comparat amb altres serveis de traducció, creus que aquesta aplicació té valor?
2. Com d'intuïtiva et sembla la manera de seleccionar el text?
3. Proposaries alguna manera alternativa de seleccionar text? (Fent servir els punters)
4. Et semblaria útil poder seleccionar una paraula concreta fent "clic" a sobre?
5. Et sembla útil que es mostri el text seleccionat en temps real mentre l'estàs triant?
6. Si hi haguessin més opcions disponibles a part de traduir i buscar, com creus que seria millor mostrar-les? I en quina posició respecte al document trobes que resultaria més útil?
7. Quines funcionalitats o característiques creus que ajudarien a millorar l'aplicació, i a donar-li més valor? (Tria totes les que vulguis)

Es pot apreciar que algunes preguntes presenten varies opcions que l'usuari pot triar, mentre que unes poques permeten a l'usuari escriure la seva pròpia resposta.

Amb les dades recollides de les enquestes podrem recollir

l'opinió dels usuaris, per tant podrem obtenir resultats del funcionament de l'aplicació, i també en podrem extreure certes conclusions preliminars, com possibles millores de futur per a aplicacions d'aquest estil. Aquests resultats es mostraran en els apartats posteriors.

En un principi, l'aplicació ha estat disponible des del dimarts 17 de maig fins al divendres 20 de maig, amb possibilitat d'ampliar aquest període si es creu necessari per a obtenir més opinions dels usuaris. Prèviament, es van agafar les dades necessàries, concretament els executables del programa i les dades dels documents i les paraules de la base de dades, i es van introduir al sistema de la biblioteca per a comprovar que el funcionament fos correcte. Un cop comprovat que l'aplicació s'executava correctament, ja es va deixar instal·lada en les dates esmentades prèviament.

Millorant el traductor de documents

Full de recollida de dades sobre l'experiència d'ús.

EDAT: _____

SEXE: _____

1. Comparat amb altres serveis de traducció, creus que aquesta aplicació té valor?

☐ Sí ☐ No

2. Com d'intuïtiva et sembla la manera de seleccionar el text?

☐ Gens intuïtiva ☐ Poc intuïtiva ☐ Intuïtiva ☐ Bastant intuïtiva ☐ Molt intuïtiva

3. Proposaries alguna manera alternativa de seleccionar text? (Fent servir els punters)

4. Et semblaria útil poder seleccionar una paraula concreta fent "clic" a sobre?

☐ Sí ☐ No

5. Et sembla útil que es mostri el text seleccionat en temps real mentre l'estàs triant?

☐ Sí ☐ No

6. Si hi haguessin més opcions disponibles a part de traduir i buscar, com creus que seria millor mostrar-les? I en quina posició respecte al document trobes que resultaria més útil?

Escriu la teva proposta a sota o, si t'és més còmode, fes un petit esbós a la part de darrere

7. Quines funcionalitats o característiques creus que ajudarien a millorar l'aplicació, i a donar-li més valor? (Tria totes les que vulguis)

- ☐ Resultats de recerca ampliat (definicions, sinònims, ...)
- ☐ Millorar els resultats obtinguts en la recerca mitjançant informació de recerca semàntica
- ☐ Interacció amb altres tipus d'elements, com imatges o gràfiques
- ☐ Poder utilitzar qualsevol document
- ☐ Poder triar idioma tant d'origen com de destí
- ☐ Autodetectar idioma del document
- ☐ Ampliar la selecció d'idiomes disponibles per a traduir
- ☐ Altres:

Figura 6. Format de l'enquesta

7 DISCUSSIÓ DELS RESULTATS

A continuació passarem a detallar els resultats obtinguts de les enquestes respostes pels usuaris. En primera instància, cal mencionar que la majoria dels usuaris veuen el concepte d'aquest tipus d'aplicació com a positiu i útil. Deixant de banda el fet de que el que van provar era un prototip amb la funcionalitat bàsica, molts usuaris van transmetre directament que una aplicació amb aquest tipus de funcionalitat era un concepte interessant i que, amb les millores corresponents, podria esdevenir una eina molt útil en l'ajuda a l'aprenentatge d'idiomes, que

era l'enfocament principal que se li va donar a l'hora de mostrar-lo als usuaris. Alguns usuaris potser no entenen a primera vista el funcionament de l'aplicació, però un cop se'ls va explicar com podien utilitzar-la van respondre positivament.

Aquesta mostra de resultats es basarà en un conjunt reduït d'enquestes, concretament 8 mostres. Aquestes mostres s'han obtingut de persones amb edats compreses entre els 27 i els 66 anys, i de les 8 mostres, la meitat l'han contestat homes i l'altra meitat dones.

7.1 Pregunta 1

7 de les 8 persones enquestades, un 87.5%, creuen que aquesta aplicació té valor respecte d'altres serveis de traducció, mentre que la restant ha deixat aquesta pregunta sense contestar.

7.2 Pregunta 2

En aquesta pregunta hi ha varietat de respostes. Sobre la manera de seleccionar el text, 4 persones (50%) creuen que és bastant intuïtiva, 2 persones (25%) creuen que és intuïtiva, 1 persona (12.5%) creu que és molt intuïtiva, i la persona restant (també 12.5%) creu que no és gens intuïtiva. Es poden veure els resultats a la Figura 7.

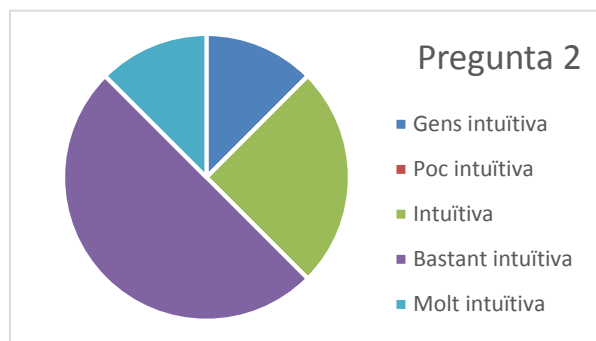


Figura 7. Gràfica amb els resultats de la pregunta 2

7.3 Pregunta 3

Una majoria dels enquestats, concretament 6 dels 8 (75%) proposen que es pugui utilitzar el text utilitzant un sol punter en comptes de definir la bounding box amb dos punters. Algunes de les propostes són:

- Dibuixar la bounding box
- Poder-la fer rectangular o circular
- Resseguir el text a escollir, com si s'estigués marcant amb un subratllador
- Arrossegar el punter, i triar totes les paraules entre el punt inicial (quan es posi el punter a la superfície) i el punt final (quan s'aixequi el punter)
- Prémer un cop el punter per definir el punt inicial de la bounding box, i tornar a prémer per definir el punt final

Les altres 2 persones (25%) no proposen cap alternativa.

7.4 Pregunta 4

A les 8 persones (100%) els seria més útil poder fer "tap" a sobre del document per a poder seleccionar una paraula individual. Cal recordar que, tot i que en un principi aquesta funcionalitat estava implementada, degut al canvi en la manera de detectar les paraules en el document a l'hora de seleccionar-les no va ser possible mantenir-la, tot i que seria possible trobar la manera de fer-ho.

7.5 Pregunta 5

7 de les 8 persones (87.5%) troben útil que es mostri el text en temps real mentre s'està fent la selecció, mentre que la persona restant (12.5%) no ho veu útil.

7.6 Pregunta 6

Aquesta pregunta només ha sigut resposta per 2 persones (25%), tot i que alguna de les respostes és bastant esclaridora. Una resposta seria que les opcions es mostressin a la part dreta del document, fet que obligaria a redistribuir els altres elements de la interfície. L'altra resposta va més enllà, i proposa que s'utilitzin menús desplegable, però que el menú estigui fixe a la part esquerra del workspace. Aquest usuari en concret va trobar que, a l'hora de seleccionar opcions del menú, el fet de que aquest estigués situat a la part superior del document podia dificultar el funcionament de l'aplicació pel simple fet de que, per haver de seleccionar una de les dues opcions, moltes vegades es posava el braç o la mà per sobre del document, fent que a vegades la càmera deixés de detectar-lo.

7.7 Pregunta 7

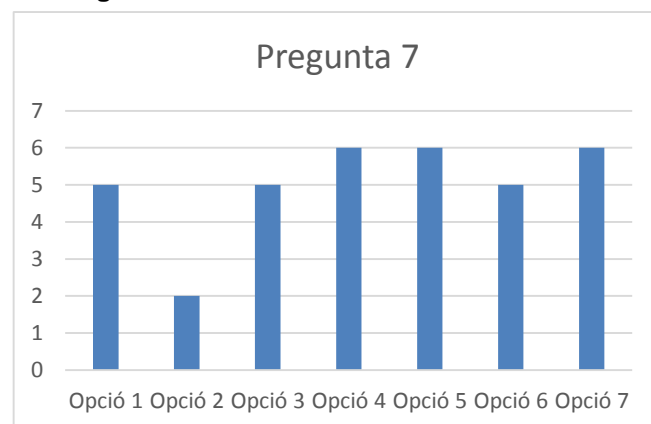


Figura 8. Gràfica amb les respostes de la pregunta 7

En aquesta pregunta s'oferien una sèrie de funcions o característiques que es podrien afegir per millorar l'aplicació, a més de donar la possibilitat de proposar-ne més. Es proposaven les següents opcions:

1. Resultats de recerca ampliat (definicions, sinònims, ...)
2. Millorar els resultats obtinguts en la recerca mitjançant informació de recerca semàntica
3. Interacció amb altres tipus d'elements, com imatges o gràfiques
4. Poder utilitzar qualsevol document
5. Poder triar idioma tant d'origen com de destí
6. Autodetectar idioma del document

7. Ampliar la selecció d'idiomes disponibles per a traduir.

La Figura 8 mostra les opcions triades pels usuaris. A més, alguns usuaris han fet les següents propostes addicionals:

- Poder seleccionar no només text del document, sinó també del resultat de la traducció
- Possibilitat de que l'aplicació ofereixi feedback auditiu. Per exemple, que es reproduïxi el text auditiu amb alguna llibreria Text To Speech, per a que així l'usuari pogués sentir la pronunciació correcta de les paraules mostrades.

8 CONCLUSIÓ

Recordem que la hipòtesi plantejada en aquest projecte és la següent: Resulta atractiva per a un usuari una aplicació desenvolupada utilitzant una interfície d'Interacció Humà-Document? Són realment funcionals aquests tipus de sistemes? La manera de validar aquesta hipòtesi ha estat utilitzar l'aplicació d'interacció Document-Humà que durant la durada d'aquest projecte s'ha estat desenvolupant, per a que un grup d'usuaris en provessin el funcionament i donessin feedback del funcionament i la utilitat de la mateixa. Aquests resultats són els que hem descrit a l'apartat anterior, obtinguts de les enquestes emplenades pels usuaris, i a través d'aquests, així com d'opinions recollides directament de boca d'alguns d'ells, en podem extreure algunes conclusions. Aquestes conclusions serien més aviat possibles millores de futur per a aplicacions d'un sistema d'Interacció Humà-Document.

Val a dir que, com ja s'ha comentat a l'apartat anterior, molts dels usuaris creuen que aquest tipus d'aplicacions poden arribar a ser útils, amb les funcionalitats degudament millorades i ampliades. Aquesta aplicació, des del punt de vista comercial, potser no tindria sentit de ser, degut a que seria probablement difícil que un usuari pogués disposar a la seva llar d'un sistema d'Interacció Humà-Document com el que s'ha utilitzat. Però sí que tindria sentit, per exemple, en un entorn enfocat a l'educació i l'aprenentatge, on el sistema fos d'accés lliure pels usuaris que desitgessin utilitzar-lo. Al realitzar la prova de l'aplicació amb usuaris, s'ha vist que l'exemple més clar d'això seria l'entorn del Library Living Lab de la Biblioteca Miquel Batllori. És un espai obert al públic, on tothom podria accedir-hi, i qualsevol que estigués utilitzant les instal·lacions podria accedir al sistema i, per exemple, utilitzar algun dels llibres en idioma estranger presents a la biblioteca per aprofitar-se de les funcionalitats que ofereix l'aplicació.

Per a això, però, s'hauria de seguir millorant aquesta aplicació fins arribar a tenir un producte que fos totalment funcional. Dels resultats de les enquestes podem extreure conclusions de cap a on es podria seguir per a millorar l'aplicació i donar-li més valor afegit. Recollint les opinions dels usuaris es van trobar alguns punts clau en que pràcticament tothom coincidia. El primer pas lògic

seria introduir a l'aplicació una llibreria OCR (Optical Character Recognition) per tal de reconèixer automàticament les lletres i paraules de qualsevol document, ja que en la mostra només es podien utilitzar els documents prèviament introduïts a la base de dades. Una de les llibreries que s'havia pensat que es podria utilitzar és Tesseract[14]. Això permetria donar total llibertat a l'usuari d'introduir qualsevol document, fins i tot de qualsevol tipus i mida. Això portaria al següent pas lògic, que seria el de donar l'opció de determinar prèviament l'idioma des del qual volem traduir, o bé utilitzar les eines d'autodetecció que la pròpia API de traducció ens proporciona. Ja s'ha vist que una gran majoria d'usuaris trobarien que les dues opcions ajudarien a millorar el funcionament de l'aplicació. Evidentment, si el que volem és donar més opcions d'ús a l'usuari, la següent millora, que està directament lligada amb les que s'acaben de comentar, seria donar la possibilitat d'escollir més idiomes als quals traduir els documents. A priori es van escollir català i castellà ja que són els idiomes que es parlen aquí, però si es vol enfocar l'aplicació a l'aprenentatge d'idiomes, tan útil es traduir des de l'idioma matern a l'idioma estranger com a l'inrevés.

En quant a millores en la selecció del text, molts dels usuaris també deien, tot i que els semblava bé la manera d'escollir el text definint una bounding box amb dos punters, que els seria més intuïtiu fer-ho amb un sol punter. Concretament, un d'ells comentava que podria ser d'una manera similar a com es selecciona text amb una tablet, o fins i tot un altre usuari comentava que podria ser com si s'estigués resseguint el text que es vol seleccionar. Una altra cosa també va quedar clara, i és el fet de que s'hauria de recuperar la funcionalitat de poder seleccionar paraules individuals amb un "tap", ja que a la pràctica si es vol seleccionar una sola paraula, amb una bounding box això presentava dificultats.

AGRAÏMENTS

Agraeixo al tutor Dimosthenis Karatzas l'ajuda i consell que m'ha donat a l'hora de fer el treball. També a Marc Castelló i Jordi Abella per a l'ajuda atorgada en aspectes tècnics en alguns punts. Finalment, vull agrair a la meua família tot el suport que m'ha donat durant tota la carrera.

BIBLIOGRAFIA

- [1] D. Karatzas, V. Poulain d'Andecy, M. Rusiñol, A. Chica and Pere-Pau Vazquez, "Human-Document Interaction systems - a new frontier for document image analysis", IAPR Int. Workshop on Document Analysis Systems, Santorini, Greece, 2016.
- [2] P. Wellner, "Interacting with paper on the digitaldesk," Communications of the ACM, vol. 36, no. 7, pp. 87-96, 1993.
- [3] N. Linder, "Luminar: a compact and kinetic projected augmented reality interface", Ph.D. dissertation, Massachusetts Institute of Technology, 2011.
- [4] P. Mistry and P. Maes, "Sixthsense: a wearable gestural interface," ACM SIGGRAPH ASIA 2009 Sketches. ACM, 2009.
- [5] ITESOFT - D3.1 Report: DoD Architecture reference guide, versió 0.8, 17/07/2014, Centre de Visió per Computador.

- [6] D. Karatzas, S. Robles, L. Gomez, "An on-line platform for ground truthing and performance evaluation of text extraction systems", Centre de Visió per Computador.
- [7] F. Guimbretière, "Paper augmented digital documents", a Proceedings of the 16th annual ACM symposium on User interface software and technology. ACM, 2003
- [8] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns, "Paper windows: interaction techniques for digital paper," a Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2005
- [9] J. Jacko, C. Stephanidis "Human-Computer Interaction: Theory and Practice", Setembre 2003 per CRC Press, ISBN 9780805849318
- [10] D. Lowe, "Object Recognition from Local Scale-Invariant Features", Computer Science Department, University of British Columbia.
- [11] K. D. Willis, T. Shiratori, and M. Mahler, "Hideout: mobile projector interaction with tangible objects and surfaces", in Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction. ACM, 2013.
- [12] C. Hurter, R. Lesbordes, C. Letondal, J.-L. Vinot, and S. Conversy, "Strip'tic: exploring augmented paper strips for air traffic controllers", in Proceedings of the International Working Conference on Advanced Visual Interfaces. ACM, 2012.
- [13] Adam Thierer. "The Internet of Things and Wearable Technology: Addressing Privacy and Security Concerns without Derailing Innovation." Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, November 2015.
- [14] R. Smith, "An Overview of the Tesseract OCR Engine," 2013 12th International Conference on Document Analysis and Recognition, pp. 629-633, Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2, 2007
- [15] F. Yergeau, "UTF-8, a transformation format of ISO 10646", Request for Comments: 3629, Network Working Group, Alis Technologies, November 2003